Correcting spatial-spectral crosstalk and chromatic aberrations in broadband line-scan spectral-domain OCT images: supplement

LE HAN[1] AND KOSTADINKA BIZHEVA[1,2,3,*] (iD)

[1] *Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[2] *Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada*
[3] *School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
*kbizheva@uwaterloo.ca

Supplementary Information for

Correcting Spatial-Spectral Crosstalk and Chromatic Aberrations in Broadband Line-Scan Spectral-Domain OCT images

Le Han[1] AND Kostadinka Bizheva [1,2,3, *]

[1] *Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[2] *Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[3] *School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

* kbizheva@uwaterloo.ca

Abstract: This document provides supplementary information for "Correcting Spatial-Spectral Crosstalk and Chromatic Aberrations in Broadband Line-Scan Spectral-Domain OCT". We show: (i) the scattering and reflecting OCT signal requires different image metrics to present their image quality; (ii) the swarm-intelligence (SI) algorithms are more robust to find the global optimum of the image metric than traditionally used gradient-based algorithms, especially in the low contrast scenario where the number and location of the image metric is unpredictable.

Section I. Factors that degrade the OCT resolution in LS SD-OCT images

1.1 Virtua interference pinhole effect

In this work, the x-direction is the scanning direction of the LS SD-OCT system, z-direction is the light propagation direction, and y-direction is the direction orthogonal to the xz-plane. Location (x,y,z) represents the center position of the optical volume under study, and (x_1,y_1,z_1) and (x_2,y_2) denote the coordinates of the imaged sample volume and the camera detection plane respectively.

Considering the light scattered back from the sample, when it reaches the camera detection plane, its electrical field can be expressed as:

$$E_{det}(x_2,y_2\,|\,x,y,z) = \int h_i(x_1,y_1,z_1)s(x_1-x,y_1-y,z_1-z)h_d(x_2-x_1,y_2-y_1,z_1)dx_1dy_1dz_1 \quad \text{(S1)}$$

where $h_i(x,y,z)$ and $h_d(x,y,z)$ are the coherent point spread functions (cPSFs) of the illuminating and detecting optics and $s(x,y,z)$ is the scattering distribution function of the sample. Here, the light rejected by the slit aperture is not taken into consideration since the slit is much wider than the beam waist. Note that this formula describes the single scattering regime case.

Since the OCT signal is generated by the interference of the electrical fields backscattered from the imaged object and reflected from the reference mirror, the reference electrical field acts as a virtual pinhole along the x-direction under the assumption that its spot size on the detection plane is the same for all wavelengths. Incorporating the virtual pinhole in the x-direction and the camera pixel size confinement in the y-direction as $p(x,y,Z)$, where Z is the axial shift of the reference focus from the detection plane, the effective electrical field backscattered from the position (x,y,z) is:

$$
\begin{aligned}
E_s^{eff}(x,y,z) &= \iint h_i(x_1,y_1,z_1)s(x_1-x,y_1-y,z_1-z)h_d(x_2-x_1,y_2-y_1,z_1)dx_1dy_1dz_1 \\
&\quad p(x_2,y_2,Z)dx_2dy_2 \\
&= \int s(x_1-x,y_1-y,z_1-z)h_i(x_1,y_1,z_1)[\int h_d(x_2-x_1,y_2-y_1,z_1) \\
&\quad p(x_2,y_2,Z)dx_2dy_2]dx_1dy_1dz_1 \\
&= s(x,y,z)\otimes[h_i(x,y,z)[h_d(x,y,z)\otimes p(x,y,Z)]]
\end{aligned}
\quad \text{(S2)}
$$

where $\otimes$ denotes convolution. Under the assumption that the pinhole is infinitely small, i.e., $p(x,y,Z) = \delta(x,z-Z)\delta(y)$, and the illuminating and detecting optical beams are underfilled such that the cPSFs are separable $h(x,y,z) = h(x,z)h'(y,z)$, E_s^{eff} can be simplified as:

$$E_s^{eff}(x,y,z) = s(x,y,z) \otimes [h_i(x,z)h_d(x,z-Z)h_i'(y,z)h_d'(y,z)] \quad \text{(S3)}$$

Here, $h_{i,d}$ and $h'_{i,d}$ refer to the cPSFs in x- and y-direction, respectively, considering that the beam waists and focal positions can be different in these two transverse directions. Since the illuminating beam radius along the y-direction is much broader than that of the detecting (420 μm vs. 1.5 μm), the effective cPSF of the scattering light can be further simplified as:

$$h_s(x,y,z) = h_i(x,z)h_d(x,z-Z)h_d'(y,z) \quad \text{(S4)}$$

It is clear from Equation (S4) that the effective detection focal plane in the xz-plane is axially shifted due to the reference signal, while it is not affected in the yz-direction.

The virtual interference pinhole acts as a confocal gate in the x-direction and reduces the necessity of deploying physical apertures; however, it results in phase loss if the focal planes are miss-aligned. For simplification, we only consider the effective cPSF of the scattering light in the xz-plane and assume the illumination and detection cPSFs are Gaussian with the same beam waist and focal position. The virtual pinhole's focal plane is shifted axially, leading to a shift in the effective detection focal position (Fig. 1B). The coherence transfer function (CTF), $H_{i,d}(k_x,z)$, which is the Fourier transform of the cPSF, can be represented as:

$$H_{i,d}(k_x,z) = \frac{\omega}{\sqrt{2}}exp\left(-\frac{\omega^2}{4}k_x^2 + i\frac{z}{2k}k_x^2\right).$$

The CTF of $h_s(x,z)$ is:

$$H_s(k_x,z) = H_i(k_x,z) \otimes H_d(k_x,z-Z) = \begin{cases} \sqrt{\frac{\pi}{2}}\omega e^{-\frac{\omega^2 - i\frac{2z}{k}}{8}k_x^2} / \sqrt{1 - i\frac{2z}{k\omega}}, & |z| \gg Z \\[2em] \sqrt{\frac{\pi}{2}}\omega e^{-\frac{\omega^2 + \frac{Z^2}{k^2\omega^2}}{8}k_x^2}, & z = \frac{Z}{2} \end{cases} \quad \text{(S5)}$$

It can be seen that when $|z| \gg Z$, $H_s(k_x,z)$ is irrelevant with the focus separation Z and its phase information, $e^{-\frac{izk_x^2}{2k}}$, is well preserved; however, the completely lost phasor at $z = Z/2$ leads to a failure of DAC.

Aligning the illumination, detection, and reference focal planes at the same axial position would be ideal such that phase of the OCT signal is phase lossless. However, it is impractical to fulfill this requirement for a broadband light source given the fact that the axial achromatic aberration is not negligible. Nevertheless, the failure of the DAC to properly compensate the monochromatic aberrations is significant only when $|Z| > z_R$, where z_R is the Rayleigh range (Fig. S1A). The beam size of the refocused cPSF $\tilde{h}_s(x,z)$ will not be distorted significantly if the separation Z of the focal planes is less than z_R for all wavelengths (Fig. S1B). The OCT images of a standard USAF 1951 resolution target were collected at different depths with large and small focal plane separation Z. A Gaussian filter is digitally applied to the interference fringes in the spectral space to reduce the affection of spatial-spectral crosstalk and chromatic aberrations, and the effective spectrum is centered at ~800 nm with a bandwidth of 50 nm. For each OCT image, sample signal $E_sE_s^*(x,y)$ and interference signal $E_rE_s^*(x,y)$ are separated, and the interference signal is corrected to the 5th order using iteration-based DAC to generate the aberration-free signal $E_r\tilde{E}_s^*$. According to group 7 element 6's local contrast in $E_sE_s^*(x,y)$ and $E_rE_s^*(x,y)$, the separation Z is about 110 μm in Fig. S1A and less than 40 μm in Fig. S1B. The difference of corrected resolutions in the x-direction is noticeable: in Fig. S1A3, all group

7's vertical bars are blurred at depth -10 μm and 10 μm; while in Fig. S1B3, all elements can be clearly resolved.
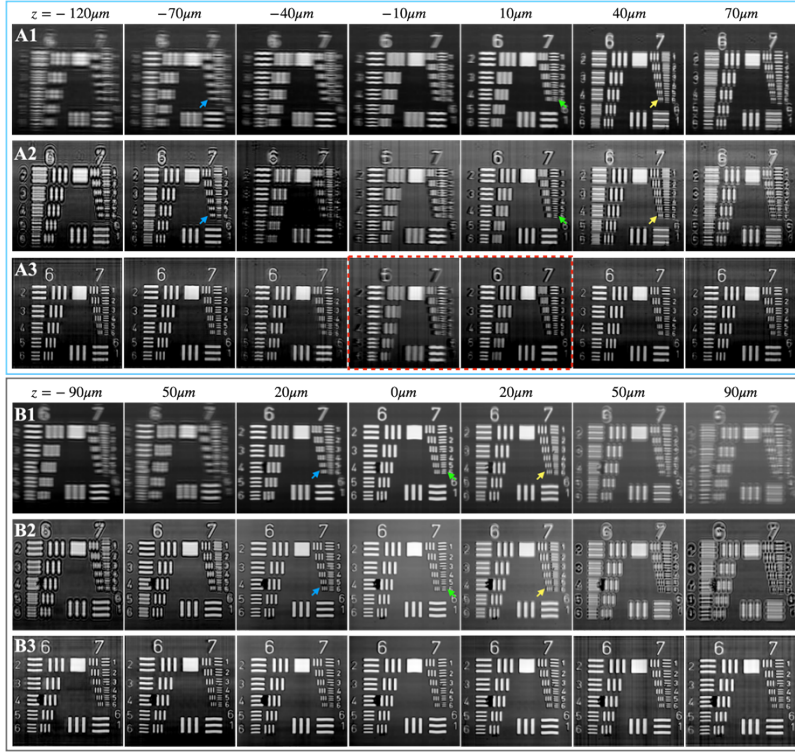

Fig. S1 DAC's performance is affected by focal planes' separation Z. (A) Log-scale enface images of groups 6 and 7 of a positive USAF 1951 target with large Z at different depths: sample signal $E_s E_s^*$ (top), interference signal $E_r E_s^*$ (middle), and aberration-corrected interference signal $E_r \tilde{E}_s^*$ (bottom). The illuminating, detecting, and reference focal planes are at depths about 40 μm (yellow arrow), 10 μm (green arrow), and -70 μm (blue arrow), respectively. At depth -10 μm and 10 μm, all group 7's vertical bars cannot be resolved by DAC (red box) since the phase information is lost. (B) Log-scale enface images of groups 6 and 7 with small Z at different depths: sample signal $E_s E_s^*$ (top), interference signal $E_r E_s^*$ (middle), and aberration-corrected interference signal $E_r \tilde{E}_s^*$ (bottom). The illumination, detection, and reference focal planes are at depth about 20 μm (yellow arrow), 0 μm (green arrow), and -20 μm (blue arrow), respectively. All elements are resolved with DAC in (B3).

Figure S2 shows the measurement of sample light intensity $I_s(z)$ and OCT axial PSF $h_z(z)$ with two different slit widths. The system's alignment is the same as in Fig. S1B. A silver mirror was used as the sample. The sample light and interference signal were recorded separately at the same depth, and each consisted of 100 repeated B-scans. I_s and h_z were averaged according to:

$$
\begin{aligned}
I_s &= \left\langle \Sigma_{p,n} I_s(y,p,n)/N \right\rangle_{y \in [Y_0-5, Y_0+5]} \\
h_z(z) &= \left\langle \Sigma_n |h_z'(y,z,n)|/N \right\rangle_{y \in [Y_0-5, Y_0+5]}
\end{aligned}
\tag{S6}
$$

where p is the pixel number along the spectral coordinate, n represents the B-scan number, $|\cdot|$ returns the absolute value, and Y_0 denotes the center of Gaussian distribution in the y-direction. Here, $h_z'(y,z,n)$ is the axially registered PSF with sub-pixel resolution; otherwise, the averaged PSF is distorted by inevitable mechanical motions. The slit widths were set as 1000 μm and

200 μm, both are much larger than the $1/e^2$ beam diameter (21 μm). Despite there being a significant difference in I_s between the two slit widths, their axial PSFs are close to each other. This result confirms that the interference acts as a confocal gate and the physical slit is just utilized to block the stray light.
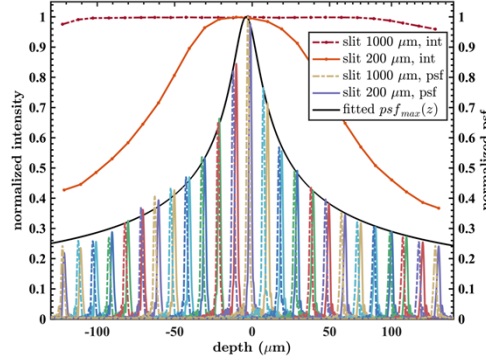

Fig. S2 Sample light intensity I_s and OCT axial PSF h_z vs. depth with different slit widths. The dashed lines represent the intensity I_s and PSF h_z measured with slit width 1000 μm. The solid lines denote the I_s and h_z measured with slit width 200 μm. The black line shows the fitted function $h_s(x = 0, z) = h_0\left(1 + \left(\frac{z - z_0}{z_R}\right)^2\right)^{-1/4}$.

Since the maximum of the ith axial PSF h_z^i is proportional to the amplitude of the effective cPSF, $h_s(x,z)$, at depth z_i, the Rayleigh range of the effective cPSF can be calculated by fitting $\left(z_i, \max\left(h_z^i\right)\right)$ with function:

$$h_s(x = 0, z) = \frac{h_0}{(1 + (\frac{z - z_0}{z_R})^2)^{\frac{1}{4}}} \tag{S7}$$

where the effective cPSF $h_s(x,z)$ is assumed to follow the electric field's distribution of a Gaussian beam. The power index $1/4$ instead of $1/2$ comes from the symmetry break of x- and y-direction. z_R is determined as 7.92 μm, corresponding to a $1/e$ beam radius of 1.42 μm of the effective cPSF.

Section II. Image metric functions

According to Parseval's theorem, Fourier transform is unitary:

$$\int |I(x,y)|^2 \, dx \, dy = \int |FT_{x,y}[I(x,y)]|^2 \, dk_x \, dk_y \tag{S8}$$

Since the DAO only applied a phase mask to the spatial frequency representation: $FT_{x,y}[I(x,y)]$ $\rightarrow FT_{x,y}[I(x,y)]\exp(i\varphi(k_x,k_y))$, the intensity integrals of the original and corrected OCT signals are the same. Therefore, the enface OCT signal is similar to an isolated system in thermodynamics: each pixel is like a particle, and the conserved pixel number and intensity integral corresponds to the isolated system exchanging no particle and energy with environment, while changing the phase mask corresponds to the system evolution. As stated in the second law of thermodynamics, the entropy of an isolated system does not decrease, it is natural to use the Shannon entropy defined in equation (5.8) as an image metric. It is rewritten here for convenience: $S_e|_{\{a_{n,i}\}} = \sum_{m=1}^{256} -p_m \log p_m$, where p_m is the probability that the normalized intensity $\hat{I}(x,y)$ is in the mth bins. Two other commonly used image metrics in aberration correction of OCT signal are:

$$S_{de}\big|_{\{a_{n,i}\}} = \sum_{x,y} -\hat{I}(x,y)\log\left(\hat{I}(x,y)\right)$$

$$S_{p}\big|_{\{a_{n,i}\}} = \sum_{x,y} \operatorname{sgn}(p-1)\hat{I}^{p}(x,y)$$

$$(S9)$$

where $\operatorname{sgn}(x)$ is the sign function. S_{de} is so-called 'entropy' or 'Shannon entropy' in the study of Synthetic-aperture radar (SAR) [1,2]; however, one should not be confused with the true Shannon entropy in information theory. The power index of S_p was set as $p = 0.9$ in this study.

The image metrics were tested with two different defocused OCT signals: one was scattered from the microspheres' phantom (Fig. S5A), the other was reflected from the resolution target (Fig. S5C). Treduce the influence of phase destruction, spatial-spectral crosstalk, and chromatic aberrations, the OCT signals were limited in a single sub-band. Only the two defocusing aberration coefficients, $a_{2,0}$ and $a_{2,2}$, were considered, and they were determined by grid searching of the global valley of $S(a_{2,2}, a_{2,0})$. The contour plots of the image metrics of the microspheres image are shown in Fig. S5B (top). From left to right, the image metrics are S_{de}, S_e, and S_p, respectively. The landscapes of the three metrics are similar except that the scales are different and S_e is noisier. The unusual number of local minima due to the noise in S_e might be related to the SNR of the signal and the number of bins used and make it challenging to apply gradient-based algorithms and SA. The grid searching result via S_e is (17.8, 23), which is slightly different from (17.4, 22,2) determined by S_{de} and S_p; although there is no noticeable difference between the three refocused images (Fig. S5B, bottom).
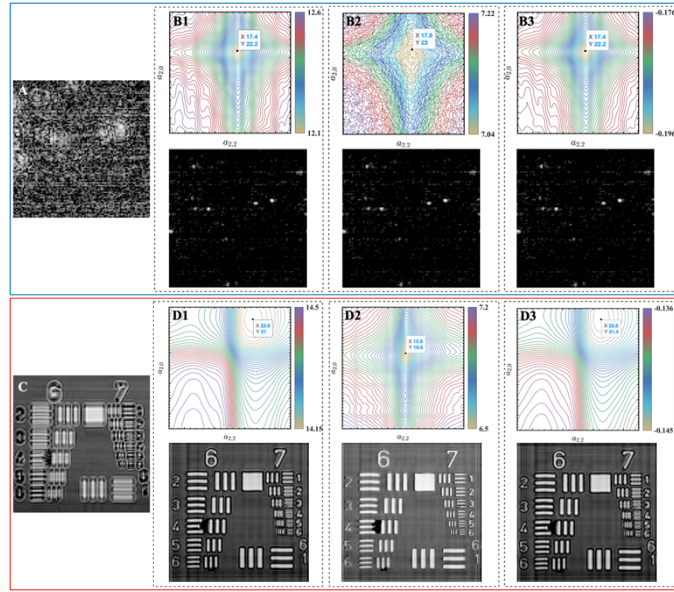

Fig. S5 Correction of the defocused scattering (top) and reflecting (bottom) OCT signals with grid search. (A) A typical log-scale enface image of the defocused microspheres. (B) Contour plot of the image metric (top) and the corresponding refocused image (bottom) using the coordinates of the global valley. From left to right, the three results correspond to S_{de}, S_e, and S_p, respectively. (C) A typical log-scale enface image of the defocused resolution target. (D) Contour plot of the image metric (top) and the corresponding refocused image (bottom) using the coordinates of the global valley. (D1) and (D3) were not fully refocused since the global minimums of S_{de} and S_p are both away from the actual defocusing coefficients.

The metrics' performances are completely different in the case of reflecting signals. The noise in S_e disappeared (Fig. S5D, middle), and the resolution target image was successfully refocused with the valley result (15.8, 19.8). However, both S_{de} and S_p failed to represent the

image quality (Fig. S5D, left and right). The global minimums of these two metrics are located at (22.6, 31) and (22.6, 31.4), respectively, whereas the actual defocusing coefficients are located at smooth slopes of the two metrics, not even local valleys. From the comparison, we can see that the Shannon entropy metric S_e can correctly represent the image quality of both the scattering and reflecting signal, although the noisy landscape in the scattering case is not suitable for some optimization algorithms. On the opposite, S_{de} and S_p work well only in the scattering case, although they are smooth and compatible with all optimization algorithms.

Section III. Optimization algorithms

We first summarize the five algorithms compared here. Two gradient-based methods are standard gradient descent algorithm and adaptive moment estimation (Adam) [3,4], while the three SI-based algorithms are accelerated particle swarm intelligence (APSO), firefly algorithm (FA), and cuckoo search (CS) [5–7]. As there is only one image data during the optimizing process in this study, there is no difference between batch gradient descent and stochastic gradient descent (SGD). In this section, we use SGD to refer to the standard gradient descent algorithm. SGD updates the parameters in the opposite direction of the metric function's gradient:

$$\vec{a}_{t+1} = \vec{a}_t - \eta \nabla S(\vec{a}_t) \tag{S10}$$

where $\vec{a}_t$ refers to the aberration coefficients' vector $\{a_j\}$ at the tth iteration, η is the learning rate, and $\nabla S(\vec{a})$ denotes the metric's gradient. Here, we use j in the subscript of the aberration coefficients instead of $\{n,i\}$ for convenience. In practice, the gradient can be calculated as:

$$\nabla_{a_j} S(\vec{a}_t) = \frac{1}{\delta}(S(\vec{a}_t + \vec{I}_j \delta) - S(\vec{a}_t)) \tag{S11}$$

where ∇_{a_j} is the partial derivative of the mth element in $\vec{a}$, $\vec{I}_j$ is a vector whose elements are all zero with the exception of one at the jth element. SGD's convergence rate is highly dependent on the chosen learning rate, and a reasonable learning rate is usually cost function dependent. In addition, it lacks the ability to escape from shallow local minima and may even get stuck at the saddle points [3].

Adam uses adaptive learning rates that are different for each parameter. The historical gradient and its square are recorded by:

$$\begin{aligned} \vec{m}_{t+1} &= \beta_1 \vec{m}_t + (1-\beta_1)\nabla S(\vec{a}_t) \\ \vec{v}_{t+1} &= \beta_2 \vec{v}_t + (1-\beta_2)(\nabla S(\vec{a}_t))^2 \end{aligned} \tag{S12}$$

where $\vec{m}_t$ and $\vec{v}_t$ are both initialized as zero. The parameter vector $\vec{a}_t$ is updated following:

$$\vec{a}_{t+1} = \vec{a}_t - \frac{\eta}{\sqrt{\hat{\vec{v}}_t} + \epsilon} \hat{\vec{m}}_t \tag{S13}$$

where $\hat{\vec{m}}_t = \vec{m}_t/(1 - \beta_1^t)$ and $\hat{\vec{v}}_t = \vec{v}_t/(1 - \beta_2^t)$. In practice, Adam can escape from the saddle points and some shallow local minima; thus, it has been one of the best overall optimization algorithms in machine learning. Nevertheless, Adam cannot guarantee a global minimum.

APSO is a variant of standard particle swarm intelligence that uses a group of agents to explore the parameters space with stochastic and deterministic components. The update rule of APSO is:

$$\vec{a}_{t+1}^{(k)} = \vec{a}_t^{(k)} + \beta(\vec{g}^* - \vec{a}_t^{(k)}) + \alpha\theta^t\vec{\epsilon}_t^{(k)} \tag{S14}$$

where (k) in the superscript refers to the kth agents, $\vec{g}^*$ denotes the historical lowest position, and $\vec{\epsilon}_t$ is a vector of random numbers drawn from a Gaussian distribution. The deterministic

update $\left(\vec{g}^* - \vec{a}_t^{(k)}\right)$ guarantees the convergence of the iteration, and its converge rate is controlled by β. The stochastic update increases the diversity of the solution searching, and its random searching region is controlled by a slowly reducing parameter $\alpha\theta^t$. The $\vec{a}_t^{(k)}$ is initialized with a random location drawn from the uniform distribution such that the agents can sample across the whole parameters space in a relatively uniform manner. The problem with APSO is that the only deterministic component may lead an aggressive converge to the current best solution $\vec{g}^*$ which can differ from the global best.

FA introduces the influence of other agents in the deterministic update, and the parameters are updated following:

$$\vec{a}_{t+1}^{(k)} = \vec{a}_t^{(k)} + \sum_{l \neq k} H(S(\vec{a}_t^{(k)}) - S(\vec{a}_t^{(l)}))\beta e^{-\gamma r_{k,l}^2}(\vec{a}_t^{(l)} - \vec{a}_t^{(k)}) + \alpha\theta^t \vec{\epsilon}_t^{(k)} \tag{S15}$$

where $H(x)$ is a Heaviside step function, and $r_{k,l}$ is the Cartesian distance of the two agents. The agent is attracted by others with lower metric values, and the attractive strength depends on the distance of the two agents. The benefit is that the agents can automatically divide into sub-groups and explore different local minima simultaneously, increasing the chance to find the global minimum.

The update rule of CS has two phases: a global random walk and a local random walk. The global random walk combines stochastic Lévy flights and a deterministic attraction to the global best location:

$$\vec{a}_{t+1}^{(k)} = \vec{a}_t^{(k)} + \alpha\tilde{L}\vec{\epsilon}_t \cdot (\vec{g}^* - \vec{a}_t^{(k)}) \tag{S16}$$

where $\tilde{L}$ refers to the step size of the Lévy flights, $\vec{\epsilon}_t$ is the direction of the Lévy flights drawn from the uniform distribution, and α is a scaling factor of the step size. Lévy flights can generate both near-field and far-field randomization. The near-field randomization explores the space like Gaussian randomization in APSO and FA, although it is biased to the global best location; the far-field randomization pushes the agents away from current locations, preventing them from getting trapped in local minima. The local random walk is performed with the possibility p_a:

$$\vec{a}_{t+1}^{(k)} = \vec{a}_t^{(k)} + \beta s H(p_a - \vec{\epsilon}_t') \cdot (\vec{a}_t^{(m)} - \vec{a}_t^{(l)}) \tag{S17}$$

where s and $\vec{\epsilon}_t'$ are a randomized scalar and vector drawn from the uniform distribution, β denotes the step size scaling factor, and $\vec{a}_t^{(m)}$ and $\vec{a}_t^{(l)}$ are two solutions of other agents that are randomly selected. CS employs a prudent update strategy that each agent adopts its update if only if the metric function drops; otherwise, the agent keeps its current location and waits for the next update. The same user-defined boundary was used for all algorithms to prevent the iteration from producing impractical solutions.

The five algorithms are tested with two scattering microspheres enface images. The two enface signals are from the same OCT tomogram but at different depths and have different aberration coefficients and SNRs. The broadened microspheres can still be recognized at the depth 75 μm away from the focal plane (Fig. S6A1), whereas the microspheres are almost overwhelmed by noise at depth 150 μm (Fig. S6A2). Data 1 and data 2 are used to refer to the enface images of the two depths in the following discussion. The refocusing coefficients, $(a_{2,2}, a_{2,0})$, are determined with the grid searching method (Fig. S6C), and the corresponding refocused images are shown in Fig. S6B. It can be seen that there are much more shallow local minima and saddle points in the metric landscape of data 2, even though just in 2 degrees of freedom. These local minima and saddle points come from the noise, and it is expected there will be more in a higher dimension.

We compared the performance of the algorithms in low order and high order aberration correction. For low order correction, the images were corrected to the 2nd order with 3 degrees of freedom; for high order correction, they were corrected to 5th order with 18 degrees of

freedom. APSO, FA, and CS employed (15, 15, 15) agents in low order correction and (40 40 20) agents in high order correction. All algorithms were performed 100 times and the image metric convergence curves are shown in Fig. S3. The grey curves show individual optimization trials, and the black line represents the mean value. The relative convergence rate and the accuracy of the five algorithms are summarized in Table S1. Here, 'd1n2' means correcting the data 1 (Fig. S6 (A1)) up to the 2nd order.
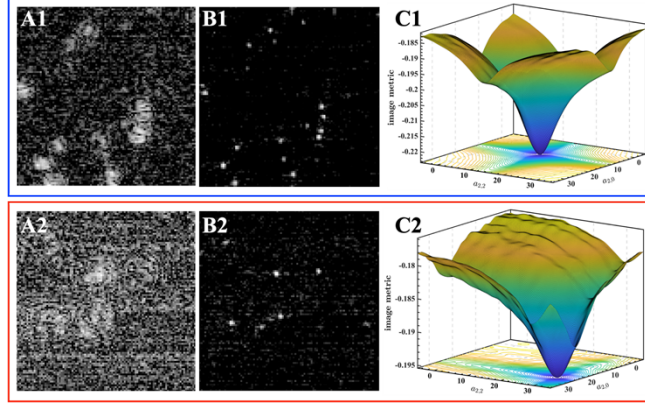

Fig. S6 OCT signals from two different depths. (A) Log-scale enface images of the original OCT signal. (B) Refocused enface images. (C) Image metric $S_p(a_{2,2}, a_{2,0})$.

The relative converge rate is inversely proportional to the function evaluation number at which the mean convergence curve reaches the critical value set as 99% minimum. If the mean curve is above the critical value, it does not have a meaningful relative convergence rate. The gradient-based evaluation numbers are equal to the iteration numbers, while the function evaluation number per iteration should be multiplied with the iteration number to determine the total evaluation number for the SI-based algorithms. The accuracy was calculated as:

$$\text{accuracy} = \frac{N_{S<(S_{min}+0.001)}}{N_{total}} \tag{S18}$$

where $N_{S<(S_{min}+0.001)}$ is the number of trials whose image metric converged below S_{min} + 0.001, and S_{min} denotes image metric's global minimum. As expected, SGD fails in all tests (Fig. S3A1-D1) except for low order correction of data 1. Only several trials found the global minimum in higher-order correction of data 1 and low order correction of data 2, and no trials succeed in high order correction of data 2. Adam, APSO, FA, and CS perform very well in low order correction of data 1 (Fig. S3A2-A5). All trials of each algorithm found the global minimum. The only difference is their convergence rate. When it comes to high order correction of data 1 (Fig. S3B2-B5), FA and CS outperform Adam and APSO in terms of accuracy. The aggressive convergence strategy contributes to the low accuracy of APSO. A considerable performance difference of the four algorithms appears in correcting data 2, which has lower SNR and more local minima. 37% and 17% trials of Adam and FA do not find the global minimum in low order correction of data 2, while nearly all trials of APSO and CS succeed (Fig. S3C2-C5). As for the high order correction of data 2, only CS maintains an acceptable accuracy above 70% (Fig. S3D2-D5).

Table. S1 Relative convergence rate and accuracy of different optimization algorithms.

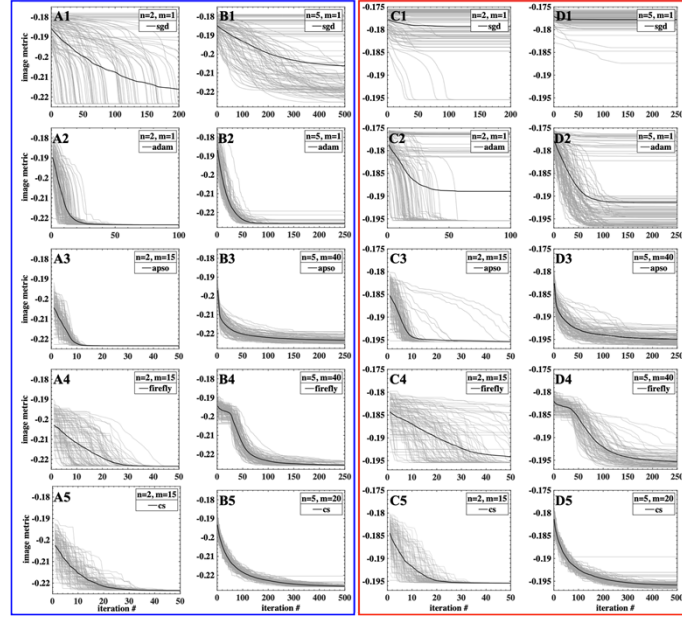| algorithm	relative convergence rate				accuracy			
	d1n2	d1n5	d2n2	d2n5	d1n2	d1n5	d2n2	d2n5
SGD	-	-	-	-	0.84	0.04	0.05	0
Adam	12	1	-	-	1	0.84	0.63	0.35
APSO	7.6	0.12	7.8	0.14	1	0.52	0.99	0.42
FA	2.8	0.23	1.7	0.14	1	0.92	0.83	0.53
CS	1.7	0.09	2.3	0.11	1	0.95	1	0.74

Fig. S3 Convergence comparison of different optimization algorithms. (A) Low order aberration correction of data 1. (B) High order aberration correction of data 1. (C) Low order aberration correction of data 2. (D) High order aberration correction of data 2. Each optimization trial is shown in grey, while the mean value is presented in black.

The accuracy defined above shows the percentage of trials that successfully find the global minimum but not the accuracy of the aberration coefficients. The first three aberration coefficients, $(a_{2,0}, a_{2,1}, a_{2,2})$, determined by different algorithms and the corresponding final image metric values are shown in Fig. S4. The abnormally large uncertainties of the three coefficients in all four test scenarios suggest that SGD cannot predict reliable results (Fig. S4A), consistent with the convergence curves in Fig. S3 (A1-D1). Adam produces acceptable mean values in correcting the aberrations in data 1, although the results in high order correction were not precise enough; however, it does not produce trustable results in data 2, whose SNR was low (Fig. S4B). The mean values of the aberration coefficients generated by APSO are decent in all four test scenarios, whereas the standard deviations in correcting high order aberrations of data 1 and data 2 are both relatively large (Fig. S4C). With the same standard, FA produced reliable results when correcting the aberrations in data 1(Fig. S4D), and CS generated the most precise results in all four cases (Fig. S4E). It only works well for low order aberration correction of images with high SNR, APSO is suitable in low order correction scenarios, FA is eligible to deal with high SNR image, and CS is the only one that works well in all four cases.
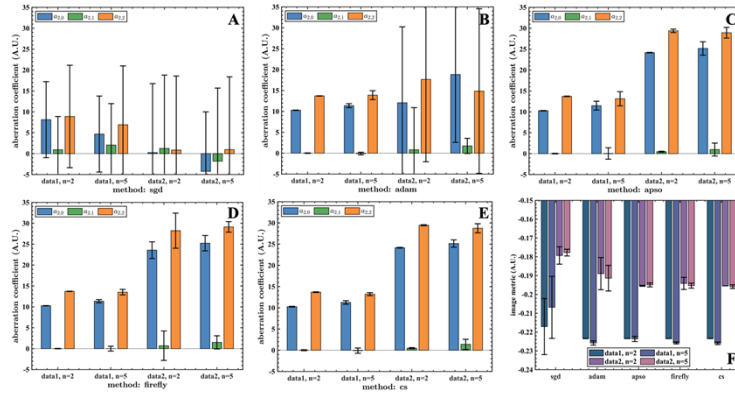
Fig. S4 Aberration coefficients determined by different optimization algorithms and the corresponding image metric values. (A-E) Mean and standard deviation of $(a_{2,0}, a_{2,1}, a_{2,2})$ determined by SGD, Adam, APSO, FA, and CS, respectively. (F) Mean and standard deviation of the final image metric values.

References

1. J. Fienup and J. Miller, "Aberration correction by maximizing generalized sharpness metrics," J Opt Soc Am 20(4), 609 (2003).
2. L. Zhang, J. Sheng, J. Duan, M. Xing, Z. Qiao, and Z. Bao, "Translational motion compensation for ISAR imaging under low SNR by minimum entropy," Eurasip J Adv Sig Pr 2013(1), 33 (2013).
3. S. Ruder, "An overview of gradient descent optimization algorithms," Arxiv (2016).
4. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Arxiv (2014).
5. X.-S. Yang, S. Deb, and S. Fong, "Accelerated particle swarm optimization and support vector machine for business optimization and application," Comm Com Inf Sc 53–66 (2011).
6. X. Yu, X. Liu, J. Gu, D. Cui, J. Wu, and L. Liu, "Depth extension and sidelobe suppression in optical coherence tomography using pupil filters," Optics Express 22(22), 26956–26966 (2014).
7. X.-S. Yang and S. Deb, "Cuckoo Search via LÉvy Flights," 2009 World Congr Nat Biologically Inspired Comput Nabic 210–214 (2009).